DOCUMENT RESUME

ED 057 740 HE 002 725

AUTHOR Wright, Lindsay G.
TITLE The Assessment of Students by Formal Examination.
INSTITUTION New Zealand Univ. Students Association, Wellington.
PUB DATE Aug 71
NOTE 37p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Evaluation; *Higher Education; *Student Evaluation; *Test Reliability; Tests

ABSTRACT

This paper presents an argument against traditional evaluation of students by examination and offers proposals for reform of the present system. Strengths and weaknesses of evaluation methods such as objective tests, use of the year's work, essay examinations, practical examinations, and oral examinations are discussed as well as the need for changes in examination conditions. (HS)

ED057740

RESEARCH OFFICE FOR THE STUDY OF HIGHER EDUCATION

ASSESSMENT OF STUDENTS BY FORMAL EXAMINATION

Lindsay G. Wright

(Education Research Officer)

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

HE002725

New Zealand University Students' Association
P.O. Box 6368, Wellington, New Zealand

August 1971

Preface

Most universities possess enough myths, both in number and variety, for the justification of almost any type of action. The greatest myth of all is that no myths govern university policies. A common assumption - a gnostic pretension - is that, of all society's institutions, the university is devoted basically and fiercely to the discovery and transmission of truth. With truth as the constant goal, who could ever doubt that university staff could be other than perfectly reasonable and highly efficient? and who, apart from staff and students, would dare to question the annual ceremonial procedures for evaluating those to whom truth or knowledge is being transmitted.

Can one question a myth without enstranging those who, as the propagators and therefore as the victims of that myth, are most committed to the myth? The main contention in this study ought to offend only the most ardent advocates of traditional university examinations - those who believe, and who will go to their graves believing, that there is one and only one way to assess exposure to university education, and that way is through an end-of-year three-hour unseen examination. If such attitudes are referred to as the endorsement of a myth it ought not to be thought that I am being selectively disparaging. All too often we forsake one myth for another. And if we were to jettison traditional university assessment techniques without either proving (or intending to test the hypothesis) that the alternative is an improvement then we are indiscernible from those who refuse to examine the alternative.

There are, I suggested above, enough myths to justify almost any type of action. If what follows is myth I would hope that academics might find, interwoven in the narrative, the occasional mundane observation that would justify the effort to read the text. Hopefully, some might come to share my suspicion that there are no university disciplines that could not vary or improve their assessment techniques.

Marian Logeman
Education Vice President
N.Z.U.S.A.

CONTENTS

	Page
Preface by Marian Logeman	
Introduction	
The Need for Assessment	2
Traditional Examinations	5
Improving Examinations	8
Objective Tests	8
Use of the year's work	10
Essay Examinations	11
Practical examinations	12
Oral Examinations	13
Changes in examination conditions	14
Miscellaneous	16
Grading students	16
Restricted and Aegrotat passes	18
Conclusion	19
Appendix: The efficient use of laboratory time in the teaching of engineering. by D.E.P. Jenkins.	20

Introduction

The first paper in this series from the Research Office for the Study of Higher Education was on The Assessment of Students by Formal Assignments and was written by Dr Michael Bassey on the basis of his experiences in teacher training.

For those who prefer reform to revolution this, the second paper, looks more at a range of possible changes in examination procedures and is a study of changes that could be implemented and what their effect might be.

While the advantages and disadvantages of continuous assessment are not, therefore, explicitly covered in this paper, it ought to be borne in mind that continuous assessment is posed as a serious alternative to examination reform. Dr Bassey's paper ought therefore to be treated as complementary to this one.

Assessment by formal examination is, in a variety of forms, an integral part of contemporary New Zealand universities. These examinations, whether in the guise of objective tests, essay, practical or oral examinations, have been subjected to considerable scrutiny over the past few years.

This particular paper is really a summary, hopefully with a few new angles, of thinking in New Zealand on examination reform and brings together a number of aspects of examining that are frequently (and perhaps misleadingly) separated in some studies.

The New Zealand University Students Association hopes that university teachers will be able to profit from the study and that it may provide useful material for those engaged in thinking about (and acting on) proposals for academic reform.

At the same time it must be clearly understood that neither the New Zealand University Students Association nor any of the constituent students' associations are necessarily committed to all the viewpoints expressed in this paper.

Lindsay G. Wright,
Education Research Officer

The Need for Assessment

1. The assessment of university students is an integral part of the whole process of university teaching and learning. This proposition, while it might be at variance with the views of some students, will be taken as self-evident in the greater part of this paper. Some considerstion however, might well be given to the aims and social functions of assessment and to the relationships between assessment and other aspects of the teaching process.

2. It could be argued that the assessment of students has at least three distinct educational aims. First, there might be value in assessing progress and attainment. Perhaps it is a human characteristic that an individual needs to be able to know from time to time both that he has improved his standards of achievement and that there are some "objective" criteria against which achievements can be measured. Secondly, students might be assessed for 'diagnostic' purposes, to reveal strengths and weaknesses, whether in particular areas or generally. It is possible that students can gain ideas of the direction in which their studies should go as a result of such feedback. The third possible aim would be that of 'prognosis' to determine which students might be capable of further study. If an assessment technique is to be a reliable guide in this latter respect it ought to be able to discriminate fairly accurately between those who could handle the study at a higher level, and those who could not. It would thus be a recognition of the imprecisionof a "prognostic" test if the subsequent performances of students could not be generally predicted from a purportedly prognostic test. A university that claimed that its assessment techniques were geared towards the three 'educational' aims referred to above would do well to ask to what extent the aims are realised.

3. Whether any social functions of assessment should exist, such that their recognition by the university would entail certain types of assessment rather than other types, is a question that could entertain the contemplative soul. Certainly it would be difficult to deny (or for that matter, to diminish) professional, public, and student demand for the certification of the academically "successful". Certification, in turn, needs to be related to some standard of competence, and assessment of competence would be a prerequisite. This point is argued well in paragraphs 8-16 of the Victoria University of Wellington Report of the Professorial Board Committee on Examining (March 1971)

4. Assessment ought, some argue, to relate to the whole teaching and learning process and, in particular, ought to be geared at ascertaining the extent to which the objectives of a course have been reached. Sometimes explicitly, but more often, implicitly, the argument goes on, university teachers do set course objectives. They set objectives insofar as they devise assessment tools which presuppose that, by the end of the university teaching year, students will be able to demonstrate their acquisition of, for example, certain skills and knowledge. If the contents of examination papers are analysed in an attempt to deduce examiner's objectives the results may show definite areas of weakness. Where this exercise

has been undertaken it has usually been found that students were predominantly required to regurgitate knowledge, to describe well-known materials or experiments, or to solve familiar problems. Less than 30 per cent of each paper required genuine thought on the part of candidates, by obliging them to attempt unfamiliar problems or by demanding the capacity to interpret or evaluate data, views, or conclusions. While academics might spontaneously question the applicability of such a judgment to their own work they may more readily recognize the laments of their colleagues who claim that too many students learn for examinations rather than because of the merit of the subject. Such a claim could suggest a divergence between what the examinations test and the attainments which lecturers believe necessary for a genuine understanding of their subject. This kind of difference most probably derives from the lecturer's failure to analyse his course objectives and to ensure that each is evaluated.

5. It should be clear by now that the word "objectives" is not being used here as a synonym for the "aim of education" or the "role of the university." The need for course objectives transcends the more speculative exercise of debating the respective merits of utilitarian or liberal education. For, whether one believes that the university should study and teach those things which are useful to society as a whole, or whether one adopts Cardinal Newmans oft-quoted and appealingly vague comment that "Liberal education makes not the Christian, not the Catholic, but the gentleman," there still remains a need to be concerned with the less exalted details of pedagogical method.

6. If then, we are not examining the long-term aims of university education, nor are we referring to the long-term aims of specific academic courses. These latter aims are certainly relevant to the whole teaching process for no academic would willingly recommend degree certification to someone who had not mastered the relevant skills and attitudes. Such long-term aims are in fact decided by academics on professional, social and philosophical grounds and provide a framework for the definition of intermediate objectives. Thus a

describe well-known materials or experiments, or to solve familiar problems. Less than 30 per cent of each paper required genuine thought on the part of candidates, by obliging them to attempt unfamiliar problems or by demanding the capacity to interpret or evaluate data, views, or conclusions. While academics might spontaneously question the applicability of such a judgment to their own work they may more readily recognize the laments of their colleagues who claim that too many students learn for examinations rather than because of the merit of the subject. Such a claim could suggest a divergence between what the examinations test and the attainments which lecturers believe necessary for a genuine understanding of their subject. This kind of difference most probably derives from the lecturer's failure to analyse his course objectives and to ensure that each is evaluated.

5. It should be clear by now that the word "objectives" is not being used here as a synonym for the "aim of education" or the "role of the university." The need for course objectives transcends the more speculative exercise of debating the respective merits of utilitarian or liberal education. For, whether one believes that the university should study and teach those things which are useful to society as a whole, or whether one adopts Cardinal Newmans oft-quoted and appealingly vague comment that "Liberal education makes not the Christian, not the Catholic, but the gentleman," there still remains a need to be concerned with the less exalted details of pedagogical method.

6. If then, we are not examining the long-term aims of university education, nor are we referring to the long-term aims of specific academic courses. These latter aims are certainly relevant to the whole teaching process for no academic would willingly recommend degree certification to someone who had not mastered the relevant skills and attitudes. Such long-term aims are in fact decided by academics on professional, social and philosophical grounds and provide a framework for the definition of intermediate objectives. Thus a medical school may quite properly say that its long-term aim is to produce good all-round general practitioners, specialists of a variety of types, and medical scientists and to ensure that certain specified or unspecified qualities of mind are inculcated. More specifically a medical school lecturer may outline a number of aims of the "to understand hypoglycaemia" variety. A clear set of course objectives would provide specific statements such as: "to list four principal causes of hypoglycaemia, to recognise the difference between acute and chronic cases, to describe the changes in the brain, to understand the mechanism, to list the minor causes, to differentiate between the lesions of acute and chronic hypoglycaemia." It is at this latter and more detailed level that a clear exposition of course objectives is related to assessment. And, while such details may be taken for granted by the expert in the field, students and those less familiar with the field will probably see the advantages of having objectives spelled out in this way.

7. Course objectives, spelled out by themselves, do lend themselves to rote learning and should be, some may well argue, moderated by less mundane objectives. The good university lecturer could spell out to students, in language suited to his own discipline, that the course objectives include a variety of objectives in the

cognitive and affective areas. The cognitive continuum, while it may start with the recall and recognition of knowledge, could include comprehension, application of information, analysis, synthesis, and evaluation. On the affective side most university lecturers will recognise their duty to try to develop in students an inner incentive to learn (motivation), an active response to stimulis so that they cover the assigned work voluntarily and with personal satisfaction, and a value such as developing a continuing determination to speak and write effectively, an appetite for what is good in literature, or the wish to achieve accurate and well developed argument in writing a report. In addition the university teacher should have as objectives that students should be able to organise their values into an internally consistent system and that, therefore, the students may be equipped to integrate their beliefs, ideas, and attitudes into a total "philosophy" or world view. Lecturers might well ponder the extent to which their assessment techniques (examinations etc) do in fact relate to all or any of the cognitive and affective objectives.

8. Assessment, then, can be regarded as having educational aims, social functions, and a close connection with course, cognitive and affective objectives. While the material in this paper that deals directly with examinations and other assessment techniques might not spell out the relationship between these techniques and the varieties of assessment procedures, the relationship should be borne constantly in mind.

Traditional Examinations.

9. Let us suppose, at the outset, the existence in New Zealand of a number of university lecturers and students who remain convinced of the efficiency of "traditional examinations." What, for the purposes of this section of this study, are they defending; why are they defending it; what are the strengths and weaknesses in their defence; and how do they counter criticisms from educationists?

10. In this study the expression "the traditional examination" will be used to refer to those techniques for the assessment of students which have most of the following characteristics.

(a) At one or more times during the academic year, and almost certainly at the end of the third term, students are obliged to attend at a scheduled place at a predetermined time to respond to a one, two, or three hour test.

(b) For the end-of-year examination at least the questions asked, whether essay type, problem solving, "objective" short-answer, or other, will not have been specified to the student directly.

(c) Within a time-limit the student must answer a set number of questions, usually without recourse to books or library facilities, and often in the presence of an invigilator other than the member of staff who set the examination.

(d) The student may, though not inevitably, never see his written script again.

(e) The greater part (e.g. more than fifty per cent) of the marks credited to a student's academic record will be based on the kind of test described above.

By "traditional examination" then we shall be meaning both the end-of-year finals and the finals moderated by results of term examinations.

11. Described by some students as a "grotesque memory orgy" and a "masquerade" the traditional examination system is still viewed however, by many students and staff as an irritating but necessary component of

Traditional Examinations.

9. Let us suppose, at the outset, the existence in New Zealand of a number of university lecturers and students who remain convinced of the efficiency of "traditional examinations." What, for the purposes of this section of this study, are they defending; why are they defending it; what are the strengths and weaknesses in their defence; and how do they counter criticisms from educationists?

10. In this study the expression "the traditional examination" will be used to refer to those techniques for the assessment of students which have most of the following characteristics.

(a) At one or more times during the academic year, and almost certainly at the end of the third term, students are obliged to attend at a scheduled place at a predetermined time to respond to a one, two, or three hour test.

(b) For the end-of-year examination at least the questions asked, whether essay type, problem solving, "objective" short-answer, or other, will not have been specified to the student directly.

(c) Within a time-limit the student must answer a set number of questions, usually without recourse to books or library facilities, and often in the presence of an invigilator other than the member of staff who set the examination.

(d) The student may, though not inevitably, never see his written script again.

(e) The greater part (e.g. more than fifty per cent) of the marks credited to a student's academic record will be based on the kind of test described above.

By "traditional examination" then we shall be meaning both the end-of-year finals and the finals moderated by results of term examinations.

11. Described by some students as a "grotesque memory orgy" and a "masquerade" The traditional examination system is still viewed however, by many students and staff as an irritating but necessary component of academic life. Those who support the traditional examinations have often attributed to them a high educational value. Preparation for examinations, they argue, trains students to deal with new material, to discriminate between the important and the unimportant, to appreciate the relevance of hitherto unrelated details, to grasp a subject as a whole and to combine parts into an organic unity, to hold knowledge ready on demand, to think for themselves. Examinations, it is further argued, have an element of reward in the student's sense of achievement (and doubtlessly an element of punishment for failure), and an element of feedback in that future directions depend on results. The feedback, some academics may say, is valuable for highlighting weaknesses or strengths, providing an incentive, giving staff an indication of their effectiveness or ineffectiveness, checking on whether teaching objectives have been achieved, serving for diagnosis and selection, and providing a return on parental or state investment. More specific defences of examinations could cover the alleged benefits of testing memory, coverage of the field, ability to relate fact and theory, ability to reason logically and to separate out points, and ability to write well - to make points clearly and succinctly. And,

where an academic is faced with detailed arguments that purport to demonstrate the advantages of alternative assessment techniques he can retreat into the assertion that while such arguments have clear prima facie cogency when applied to other subjects the contingencies of his own subject make traditional examinations either desirable or necessary.

12. The defences of examinations, including defences not cited above, do need to be studied carefully. While such a study will not form part of this paper it must be said that the defences of examinations are frequently relevant also to defences of non-examination assessment procedures. That is, assuming the necessity for some of the things that exams are said to do the question is not simply "do examinations do these things?" Rather, one should ask which form of assessment does most of these things the most successfully. In this context it is certainly appropriate to discuss questions of the validity and reliability of differing assessment techniques.

13. Most of the studies of the validity and reliability of assessment techniques seem to concentrate on examinations. The validity of an examination is, briefly, the extent to which it measures what it is, in fact, supposed to measure. Academics are prone to assert, though rarely to attempt to prove, that the exams they set measure one, some, or all of the things that are cited in para 11. Statistically obsessed exam advocates point to the highish correlations between results in the exams they set and the results in I.Q. tests, secondary school exams, and the impressions of academics themselves. These correlations, they might argue, show that whatever is being assessed in I.Q. tests and other exams is certainly related to what is being assessed in traditional university examinations.

14. The problem of examination reliability is well known, and has been attributed to a number of factors. First, student performance at any specific time may suffer as a result of physical or mental discomfort. Secondly, the examination set may not be adequate for sampling the candidates knowledge. Generally speaking, the longer the examination and the greater the number of questions to

where an academic is faced with detailed arguments that purport to demonstrate the advantages of alternative assessment techniques he can retreat into the assertion that while such arguments have clear prima facie cogency when applied to other subjects the contingencies of his own subject make traditional examinations either desirable or necessary.

12. The defences of examinations, including defences not cited above, do need to be studied carefully. While such a study will not form part of this paper it must be said that the defences of examinations are frequently relevant also to defences of non-examination assessment procedures. That is, assuming the necessity for some of the things that exams are said to do the question is not simply "do examinations do these things?" Rather, one should ask which form of assessment does most of these things the most successfully. In this context it is certainly appropriate to discuss questions of the validity and reliability of differing assessment techniques.

13. Most of the studies of the validity and reliability of assessment techniques seem to concentrate on examinations. The validity of an examination is, briefly, the extent to which it measures what it is, in fact, supposed to measure. Academics are prone to assert, though rarely to attempt to prove, that the exams they set measure one, some, or all of the things that are cited in para 11. Statistically obsessed exam advocates point to the highish correlations between results in the exams they set and the results in I.Q. tests, secondary school exams, and the impressions of academics themselves. These correlations, they might argue, show that whatever is being assessed in I.Q. tests and other exams is certainly related to what is being assessed in traditional university examinations.

14. The problem of examination reliability is well known, and has been attributed to a number of factors. First, student performance at any specific time may suffer as a result of physical or mental discomfort. Secondly, the examination set may not be adequate for sampling the candidates knowledge. Generally speaking, the longer the examination and the greater the number of questions to be answered, the more reliable will be the score. One usually needs, according to some authorities, at least four 2-hour papers for the reliability coefficient to reach 0.90 (1) This estimate of course applies to the traditional examination system. Thirdly, there may be inconsistencies in standards of marking adopted by different examiners. If a grade of, say First Class or "A" means the grade obtained by 10 per cent of the candidates in one department or university, and the grade obtained by 5 per cent of the candidates in another department, then either the same grade means different things to different departments or the quality of students in the departments varies greatly. Fourthly, there may often be different opinions among examiners about the relative value of students' answers. Most academics will be well aware of the results of experiments in which the same scripts have been assessed by a

(1) Lovell, K. "Examinations and Marking," University Teaching in Transition, Oliver and Boyd, 1968, p137.

number of independent people and the resulting marks have varied quite considerably. It we put reliability in its statistical context we find that at a high 0.9 reliability, a remarking would give 68 per cent of the students no more than 3 marks out of 100 either way, in 28 per cent marks would vary between 3 and 6 either way, and in 4% of the cases they will vary by more than 6 marks. At 0.8 reliability 68% will vary by 4 marks, 28% by between 4 and 8, and in 4% of the cases by more than 8 marks.

15. Some writings on examinations refer to a problem known as examination anxiety. While some degree of anxiety would be normal for most students who face examinations there are clear documented cases of examination stress that amount to panic and phobic anxiety. For such students examinations may represent not a circumscribed and potentially manageable challange, but an overwhelming threat and a potential failure to the total person. The existence of very small numbers of students suffering from examination phobia may justifiably be relevant to the operations of student health services but does not entail the conclusion that examinations should be abolished. Such anxieties are, we are told by student counsellors, capable often of being reduced to manageable proportions. (1)

16. It would be difficult in the

number of independent people and the resulting marks have varied quite considerably. It we put reliability in its statistical context we find that at a high 0.9 reliability, a remarking would give 68 per cent of the students no more than 3 marks out of 100 either way, in 28 per cent marks would vary between 3 and 6 either way, and in 4% of the cases they will vary by more than 6 marks. At 0.8 reliability 68% will vary by 4 marks, 28% by between 4 and 8, and in 4% of the cases by more than 8 marks.

15. Some writings on examinations refer to a problem known as examination anxiety. While some degree of anxiety would be normal for most students who face examinations there are clear documented cases of examination stress that amount to panic and phobic anxiety. For such students examinations may represent not a circumscribed and potentially manageable challange, but an overwhelming threat and a potential failure to the total person. The existence of very small numbers of students suffering from examination phobia may justifiably be relevant to the operations of student health services but does not entail the conclusion that examinations should be abolished. Such anxieties are, we are told by student counsellors, capable often of being reduced to manageable proportions. (1)

16. It would be difficult in the matter of assertion and counter assertion to say that evidence to hand, provides a sound justification either for or against the retention of the end-of-year examination. Nonetheless, the reliability of traditional examinations is highly suspect (in spite of the protestations of exam advocates); it is doubtful that course objectives are measured well by examinations currently set; and given a clear statement of course, cognitive and affective objectives it is likely that alternative assessment procedures might often be preferable. There is, it seems, little to counter, and much to support the proposition that traditional examinations serve little purpose, serve that purpose poorly, and could well be abolished.

(1) Jon Frederick. "The Management of Students experiencing acute examination anxiety." Paper to Conference of ANZSHA 1970.

Improving Examinations

17. There are, of course, those who admit the limitations of traditional examinations and who believe that the solution is not to abolish examinations but to improve them. Of the variety of improvements proposed we shall look at a few which have been tried at some university institution and which are consistent with the retention of all of the best (and possibly many of the worst) features of traditional examinations.

Objective tests

18. Objective tests, developed largely as a reaction to examination unreliability, are so denoted because of their scoring procedure. Advocates of "objective tests" may argue that they yield objective data; are reliable insofar as a test readministration will yield closely similar results; discriminate more accurately the finer shades of knowledge between candidates; eliminate the advantage gained by the speedy writer; are more easily tested for validity; and can give prompt, detailed and descriptive feedback both to the examiner and the examinee.

19. Writing questions for an objective test is a creative task, and a demanding one. The work is slow - a competent exam writer may do well to construct between five and ten good items per day. Questions must be unambiguous, alternatives must be plausible, the items must elicit responses that will indicate the level of a student's understanding or knowledge. Several types of objective tests can be used:

(a) Completion or Short Answer Items:

Where a student is asked to supply rather than select a response. This format is probably most often used for computation problems where students are required

17. There are, of course, those who admit the limitations of traditional examinations and who believe that the solution is not to abolish examinations but to improve them. Of the variety of improvements proposed we shall look at a few which have been tried at some university institution and which are consistent with the retention of all of the best (and possibly many of the worst) features of traditional examinations.

Objective tests

18. Objective tests, developed largely as a reaction to examination unreliability, are so denoted because of their scoring procedure. Advocates of "objective tests" may argue that they yield objective data; are reliable insofar as a test readministration will yield closely similar results; discriminate more accurately the finer shades of knowledge between candidates; eliminate the advantage gained by the speedy writer; are more easily tested for validity; and can give prompt, detailed and descriptive feedback both to the examiner and the examinee.

19. Writing questions for an objective test is a creative task, and a demanding one. The work is slow - a competent exam writer may do well to construct between five and ten good items per day. Questions must be unambiguous, alternatives must be plausible, the items must elicit responses that will indicate the level of a student's understanding or knowledge. Several types of objective tests can be used:

(a) Completion or Short Answer Items:

Where a student is asked to supply rather than select a response. This format is probably most often used for computation problems where students are required to generate specific solutions. It is also useful in assessing the students' knowledge of "barebones" factual material. Short Answer techniques are probably best used where the intent is to test information recall, where computation is required, or where a recognition question would make an answer obvious.

(b) Recognition and selection items:

Where a student is asked to make appropriate discriminations. While this format denies to students the latitude to redefine a problem, to offer partially correct answers, or to introduce irrelevancy it has the disadvantage that recognising a right from a wrong answer requires less thorough learning than recalling specific information.

(c) True-False items:

A format relied on by lecturers seeking an easily constructed test. Few statements however, are unquestionably true or false. And many that are do not measure the abstract and inferential learning objectives that are part of a university course. With a 50-50 chance of a correct guess for each question a test needs a large number of items then a multiple-choice test to maintain its discriminating power. True-false tests, while

appealing to those with little experience of test construction, should be used with great caution.

(d) Multiple-Choice Tests:

Well constructed multiple-choice tests are capable of measuring most educational objectives, whether the ability to solve problems, recommend actions and make predictions, or the acquisition of knowledge, understanding, and judgement. In view of a not uncommon view that objective tests require only a superficial knowledge of simple facts some more details on multiple-choice tests may be helpful.

20. The major drawback of objective tests is that bad items are so simple to construct. Both questions and answers must avoid clues which assist the uninformed student to answer correctly without really understanding the material. It is important to make sure that correct and incorrect answers are of the same length, structure, technical sophistication and grammatical form or else they will stick out as incongruous. For example there is little merit in an item such as:

"Freud was a - 1. Marine biologist
2. Psychoanalyst
3. Ethologist
4. Historian"

Similarly a question such as "what is the best form of treatment for an elderly patient with thyrotoxicosis and severe heart failure?" poses the problem as to whether the treatment is really for the heart failure or the thyrotoxicosis. The question, as a preface to a multiple-choice item, might read better as "what is the best form of treatment for thyrotoxicosis in an elderly patient

appealing to those with little experience of test construction, should be used with great caution.

(d) Multiple-Choice Tests:

Well constructed multiple-choice tests are capable of measuring most educational objectives, whether the ability to solve problems, recommend actions and make predictions, or the acquisition of knowledge, understanding, and judgement. In view of a not uncommon view that objective tests require only a superficial knowledge of simple facts some more details on multiple-choice tests may be helpful.

20. The major drawback of objective tests is that bad items are so simple to construct. Both questions and answers must avoid clues which assist the uninformed student to answer correctly without really understanding the material. It is important to make sure that correct and incorrect answers are of the same length, structure, technical sophistication and grammatical form or else they will stick out as incongruous. For example there is little merit in an item such as:

"Freud was a - 1. Marine biologist
2. Psychoanalyst
3. Ethologist
4. Historian"

Similarly a question such as "what is the best form of treatment for an elderly patient with thyrotoxicosis and severe heart failure?" poses the problem as to whether the treatment is really for the heart failure or the thyrotoxicosis. The question, as a preface to a multiple-choice item, might read better as "what is the best form of treatment for thyrotoxicosis in an elderly patient with heart failure?" The response alternatives may then contain elements relating to the treatment of both disorders. Some multiple-choice items may simply test knowledge, some may test ability to argue logically, and items like the following can do both:

"Assume that over a ten-year period consumer purchases had dropped from 75 per cent to 65 per cent of the GNP, and that investment and net exports had remained the same, proportionately. On the basis of this information alone, we can be certain that:

1. Total expenditure had declined, percentage-wise
2. The size of the GNP had declined
3. Government spending had increased, percentage-wise
4. The size of disposable income had declined
5. Personal savings had increased, percentage-wise.

The correct answer is (3).

21. A careful examiner may; having devised a multiple-choice test that tests factual recall, data interpretation, and problem solution, then pretest the examination on all of his departmental colleagues to check for ambiguities and errors and to establish some sort of standard. In particular, the examiner should attempt to ensure that some of the questions are clearly discriminatory so that the better students are able to be distinguished from the dull or mediocre. Once a lecturer has set an objective test he must then score the results, and must do so consistently. Thus for a nine-part question with three correct answers an examiner could:

1. award points for each correct answer

2. award points for each correct answer and deduct points for each wrong answer

3. award points for each correct answer identified and award points for each incorrect answer which has been identified as incorrect.

This last approach for a nine-part question can give a score range of 19 where three right answers give plus 3, six wrong answers ignored give plus 6, ending up with plus 9. In the opposite way one can score minus 9. The multiple-choice approach may then, provide the scope for a wide spread of marks, and, if used with computer marking, can save much staff time after examinations.

22. It may be argued that the above account of objective tests is too optimistic in tone and insufficiently attentive to problems of test construction for staff and problems of examination stress when such tests are used frequently. Further, university staff in New Zealand may be only too well aware of the poor test construction in some School Certificate papers using this approach. It may well be that many reservations felt by staff, particularly in the humanities and the social sciences, stem from their personal limitations rather than from problems inherent in objective testing. The rationalisations that emerge to justify staff reservations may be as sophisticated as they are fanciful and staff could do well to assess the basis of whatever doubts they have.

Use of the Year's Work

23. A second approach to the problem of improving the

2. award points for each correct answer and deduct points for each wrong answer

3. award points for each correct answer identified and award points for each incorrect answer which has been identified as incorrect.

This last approach for a nine-part question can give a score range of 19 where three right answers give plus 3, six wrong answers ignored give plus 6, ending up with plus 9. In the opposite way one can score minus 9. The multiple-choice approach may then, provide the scope for a wide spread of marks, and, if used with computer marking, can save much staff time after examinations.

22. It may be argued that the above account of objective tests is too optimistic in tone and insufficiently attentive to problems of test construction for staff and problems of examination stress when such tests are used frequently. Further, university staff in New Zealand may be only too well aware of the poor test construction in some School Certificate papers using this approach. It may well be that many reservations felt by staff, particularly in the humanities and the social sciences, stem from their personal limitations rather than from problems inherent in objective testing. The rationalisations that emerge to justify staff reservations may be as sophisticated as they are fanciful and staff could do well to assess the basis of whatever doubts they have.

Use of the Year's Work

23. A second approach to the problem of improving the traditional examination system is to "use the year's work in the final assessment." This approach may bypass the debate about assessment procedures and simply deal with the desirability of grading students by counting final examinations as worth say 75 per cent of the final marks while the year's work is graded out of 25 per cent. A number of New Zealand universities have emphasised that this approach is followed by a number of departments in their institution. The varieties of assessment procedures and the differences in the proportioning of marks between finals and the year's work makes "use of the year's work" a convenient but question-begging approach to the whole problem of improving assessment procedures. For, while it is of interest to note that departments either do, or plan to, allocate marks for the year's work they may well do so by varying their assessment procedures not one whit. For example, a department that allocates 30 per cent of the marks for the final grade on the basis of a traditionally styled examination set in the second term, and 70 per cent on the basis of a final examination could only marginally be said to consider the year's work. Similarly a Department that allocated 25 per cent of the final marks for tests, essays and practical examinations throughout the year is still relying largely on final examinations for the final gradings.

24. Where New Zealand university departments have introduced a 25/75 per cent distinction between "the year's work" and final examinations they have frequently ascribed to their procedures the merits and demerits of assessment systems that use no examinations but only the normal assignment work. Academic's can argue that such an allocation promotes more steady work, a better coverage of the year's work, improved academic standards, and a

reduction of the element of chance. On the other hand, those who wish the year's work to be free from assessment, can argue that even a, say 25/75 per cent division, re-creates an undesirable pre-university learning situation, increases demands on staff time, increases the chance of plagiarism, penalises those who work best under concentrated pressure, and limits the freedom of students to plan their own work. Some academics for whom examinations are the essence of assessment, envisage the grim spectacle of a steeplechase system of examinations. It could, it would seem be fairly argued that the "use of the year's work" does not, necessarily, but very well could, entail an improvement on the traditional examination system. In this context the "could" is contingent on staff flexibility.

Essay Examinations

25. Essay examinations have been the standard fare for university students for a long time, and many of the criticisms of traditional examining relate particularly to the limitations of subjective marking that are a feature of most essay examinations. It is, in fact, the difficulties in scoring, rather than the form of questions or their purpose, that have been most used to pinpoint weakness in the essay style examination. At the same time there are two general schools of thought about the relevance of the form of essay exams. One school argues that the essay-type question tests ability to write discursively and lucidly and shows whether or not a candidate can [illegible]

reduction of the element of chance. On the other hand, those who wish the year's work to be free from assessment, can argue that even a, say 25/75 per cent division, re-creates an undesirable pre-university learning situation, increases demands on staff time, increases the chance of plagiarism, penalises those who work best under concentrated pressure, and limits the freedom of students to plan their own work. Some academics for whom examinations are the essence of assessment, envisage the grim spectacle of a steeplechase system of examinations. It could, it would seem be fairly argued that the "use of the year's work" does not, necessarily, but very well could, entail an improvement on the traditional examination system. In this context the "could" is contingent on staff flexibility.

Essay Examinations

25. Essay examinations have been the standard fare for university students for a long time, and many of the criticisms of traditional examining relate particularly to the limitations of subjective marking that are a feature of most essay examinations. It is, in fact, the difficulties in scoring, rather than the form of questions or their purpose, that have been most used to pinpoint weakness in the essay style examination. At the same time there are two general schools of thought about the relevance of the form of essay exams. One school argues that the essay-type question tests ability to write discursively and lucidly and shows whether or not a candidate can synthesise and generalise. The other school suggests that the usual examination, with three to five questions to be discussed in three hours, tests a very specialised kind of ability, and puts a premium on speed and superficiality. It limits the student to one mode of expression, and often, so the argument continues, provides a more sensitive measure of writing speed and resistance to anxiety than of educational achievement. It should be noted that the merits proposed for the essay exam are also often attributed to essays done under non-examination conditions, whereas the demerits noted by the second school above relate specifically to examination conditions and do not contradict the merits attributed to non-examination essays. Advocates of the essay examination do, of course, raise questions relating to the feasability of proper assessment of non-examination essays but such questions are generally capable of some answer. (See, for example, Michael Bassey's Assessment by Formal Assignments, NZUSA, 1971)

26. Potential troubles with essay examinations may, some suggest, be minimised if questions are geared at a reasonable level, are worded clearly and unambiguously, are specific on the scope and length of answers required, and adequately sample instructional objectives. In marking, examiners could mark each paper without knowing who wrote it (often difficult), grade all answers to one question at one sitting, randomize the sequence in which scripts are read, and judge each essay question analytically by assigning points to various sections. Whether these mechanical procedures provide any safeguards is a moot point - they might however help to improve essay examinations. For example, an essay in answer to the question "Describe the origins of World War I" is both more difficult to answer pointedly and to mark than "Explain why each of the great European powers - Austria, Germany, France and Great Britain - must bear some of the responsibility for the events which led immediately to the outbreak of World War I." On balance, it might be said

that the essay appears capable of making a valuable contribution to student thinking and student capacity to arrange ideas, but a cumulative assessment of students' essays over a period of time might well be more reliable than the results of essay examinations. On no account, however, if essay examinations are to be set, should they be set at the end of a course in which students have written no essays at all. If essay writing is not relevant to achieving the course objectives then it has no place in assessing the extent to which these objectives have been reached. Conversely if an essay examination has to be set at the end of a given course then, since that course contains objectives best measured, in part at least, by essay work there should be essay work during the course itself.

Practical Examinations

27. The assessment of practical skills or experimental ability is an aspect of university teaching that rarely attracts the attention of educationists. If a university teacher looks for a single publication to assist him in setting better problem papers, assessing experimental work, or evaluating practical skills more objectively, he will probably look in vain. Yet for many graduate scientists, and most dentists, doctors, horticulturists, engineers, architects, etc, the professional skills are probably as important as the theoretical base on which they are built. While "practical examinations" may not form part of each aspect of the sciences and professional disciplines, they

that the essay appears capable of making a valuable contribution to student thinking and student capacity to arrange ideas, but a cumulative assessment of students' essays over a period of time might well be more reliable than the results of essay examinations. On no account, however, if essay examinations are to be set, should they be set at the end of a course in which students have written no essays at all. If essay writing is not relevant to achieving the course objectives then it has no place in assessing the extent to which these objectives have been reached. Conversely if an essay examination has to be set at the end of a given course then, since that course contains objectives best measured, in part at least, by essay work there should be essay work during the course itself.

Practical Examinations

27. The assessment of practical skills or experimental ability is an aspect of university teaching that rarely attracts the attention of educationists. If a university teacher looks for a single publication to assist him in setting better problem papers, assessing experimental work, or evaluating practical skills more objectively, he will probably look in vain. Yet for many graduate scientists, and most dentists, doctors, horticulturists, engineers, architects, etc, the professional skills are probably as important as the theoretical base on which they are built. While "practical examinations" may not form part of each aspect of the sciences and professional disciplines, they are still present in one form or another in many courses. In laboratories, workshops, drawing offices, studios, farms and hospitals university students are dissecting organisms, building machinery, designing buildings, performing on musical instruments, assessing the merits of a new crop, and interviewing disturbed patients. And, since these activities form part of the course and are undertaken presumably with some course objective in mind, they are rightly assessed.

28. With such a wide range of practical skills involved in university courses the following comments may have a bearing only on certain types of courses. For students with regular laboratory classes it might be argued that the objectives of practical classes are to teach the skill of experimentation, the value and importance of care and precision in experimental research, the design of experiments and the scope and limitations of the scientific method. Some academics might believe that students, through their own observations and measurements will see that the laws and theories they have met in their lectures can be, in part at least, confirmed in practice. The latter belief is consistent with assessment procedures for practical work that are tests of verbal information and application. For the former objectives, however, it is not at all obvious that linguistically coded information produced under examination conditions can be sensibly used in assessing a student's exposure to the segment of experience to which such information might refer. This point is well taken by many university departments and the practical examination may then involve the student in the attempt to "do an experiment" (or experiments) under close observation at a time and date specified for examination purposes.

29. This approach to the assessment of practical work, particularly where it involves a "rerun" of one or more "experiments" previously attempted as part of the course

may test manual dexterity, the ability to communicate, the extent to which facts and knowledge of procedure have been assimilated, and ability to write a report succinctly, but it provides little scope for assessing creative ability. Perhaps our practical classes do not justify any other assessment procedures. If that is so then lecturers may well have to ask themselves whether they wish to retain practical work for the purpose of illustrating learned work or to enable students to sample scientific behaviour. If academics are concerned with the latter purpose then assessment procedures may change accordingly. For example, in most practical work situations there would appear to be little or no reason why the assessment of practical work should not be done in the laboratory at the time when normal laboratory work is being done. To set a 'practical examination' when continuous assessment is either being done (or is possible) is to add unnecessarily to the course demands.

30. For those concerned with the assessment of laboratory work it may help to consider the assessment of practical work in the context of a course in which the laboratory is used for more than the common activity of performing set tasks and writing reports accordingly. A description of such a course is appended to this paper (Appendix I: The Efficient Use of Laboratory time in the Teaching of Engineering) and outlines a programme of experimental work with certain specified objectives. University staff should consider whether or not the objectives mentioned are ones they agree with, whether these objectives are currently met in their own practical courses, whether the programme outlined is consistent with their own thinking on experimental work, and whether or not such work can be assessed outside the normal laboratory setting. By posing such questions university teachers might find it difficult to avoid the conclusion that practical examinations are a poor substitute for assessment and that an enquiry oriented programme is of more value than the replication of classical and standard experiments.

Oral Examinations

31. By long standing tradition oral examinations are most frequently used in medicine, dentistry, and foreign languages. Like traditional written examinations they are reputed to provoke anxiety and to be severely limited in reliability and in the depth of evaluation which they provide. In discussing the question of oral examinations the Victoria University Report of the Professorial Board Committee on Examining (March 1971, paras 57-61) asserted that New Zealand university students are unfamiliar with the technique of "oral exposition." The Committee suggested that this unfamiliarity constituted "another argument against the use of oral examining." It may very well be true that the possibility of reliable and valid assessment by oral interrogation or even by discussion should be treated sceptically. Nonetheless, if university teachers see skill in verbal communication as an objective of university teaching they may have some basis for an attempt to assess such skills. If such skills are foreign to New Zealand students perhaps they could be introduced to them more systematically than at present. This is not to say that oral examinations might be the ideal way of assessing verbal skills - for it is in the group

discussion, tutorial and laboratory context that academic staff might best assess. However, where the examination context can be proved to be the best place for testing oral proficiency, some steps can be taken to marginally improve test objectivity. First, the examiner or examiners could compile as specific a list as possible of the areas of knowledge and skill they intend to evaluate. Secondly a permanent record could be obtained on audio or videotape so that the discussion can be examined at leisure. Thirdly, for clinical examinations in medical subjects the results of oral examination could be checked against the record of the student's performance in ward rounds, outpatients clinics, etc.

Changes in Examination Conditions

32. A number of small modifications to the traditional examination procedures have been proposed from time to time, occasionally with arguments for and against the proposal included, but rarely with details of the objectives that are likely to undergird the proposals. First, some examiners have suggested open-book examinations in which students are free to take text and reference books into an examination room. It is probable that most of the alleged merits of essay examinations would be retained equally as well if such a procedure were adopted. With poorly structured questions this system might encourage the regurgitation of text book material but could equally easily be said to test ability in the use of reference materials. Secondly, examiners could provide students with specific information about the examination questions at a reasonable time before the exam itself. Advocates of this approach might argue that the resulting lack of need for "question spotting" or complete course coverage will enable students to reveal their ability to study in depth, and may reduce nervous tension a little. Certainly it is doubtful that the provision of information would in any way alter student dependence on memory or writing skill. Those who may currently benefit from more inspired question spotting would have no clear advantage, and those who do not trust their "question spotting" ability would be relieved of the disadvantages of superficially scanning all their material before each examination. Thirdly, by setting a single-question unseen examination paper a student may be enabled to explore one topic in greater depth with an emphasis on thought and synthesis of ideas rather than on speed of writing. Most of advantages and disadvantages of the normal three hour paper are retained in this proposal except for the drawback of the superficiality imposed by shorter questions.

33. Where a university department is unwilling to lose its dependence on formal examinations it might well consider the possibility of varying the type of examination set in line with changes similar to those described in the preceeding paragraph, or by the use of "take-hone" examinations. At the same time it might well be of value to discuss whether or not there are any intrinsic merits in holding examinations of three hours duration. The Victoria University Report on Examining already referred to, argues at some length in favour of one - or two-hour examinations (paras 35 and 36). The Report asserts, inter alia, that a one-hour paper can adequately sample

a large amount of presumed knowledge; that three-hour examinations confuse assessment with tenacity; that more frequent shorter examinations can validly test as capably as one three hour final examination; and that more frequent shorter examinations could reduce the need for the "sterile regimentation" of "finals." Such claims if acceptable, could help justify a reduction in staff reliance on the end-of-year three-hour unseen examination..

Miscellaneous

34. Assessment procedures must, it can reasonably be argued, form a part of the total context of a university system. This assertion, though not without its detractors, does suggest that changes in other aspects of a university might entail the need for changes in assessment procedures and, conversely, changes in assessment procedures might entail changes in other aspects of a university. In this section of our study of assessment by formal examinations we shall look at the questions of the value of grades, and all special types of passes.

35. Grading Students

When the examination rooms empty out in mid-November and when the academic staff of our university have allocated the marks to each examination script, undergraduate students rush for the newspapers to see whether they have been awarded an A, B or C pass or a D or E failure. For many students enrolled in "intermediate" courses these grades are the basis on which they will be accepted or rejected for enrolment in a professional school. And, for students completing a Masters or Honours degree the "First Class Honours" label may be the admission ticket to an academic career or a post-graduate scholarship. While the attack on grading has not yet reached New Zealand universities to any significant degree it might be of interest to note the kind of arguments proposed by the advocates of grade abolition overseas.

36. At the outset those who seek to abolish grades might argue

34. Assessment procedures must, it can reasonably be argued, form a part of the total context of a university system. This assertion, though not without its detractors, does suggest that changes in other aspects of a university might entail the need for changes in assessment procedures and, conversely, changes in assessment procedures might entail changes in other aspects of a university. In this section of our study of assessment by formal examinations we shall look at the questions of the value of grades, and all special types of passes.

35. Grading Students

When the examination rooms empty out in mid-November and when the academic staff of our university have allocated the marks to each examination script, undergraduate students rush for the newspapers to see whether they have been awarded an A, B or C pass or a D or E failure. For many students enrolled in "intermediate" courses these grades are the basis on which they will be accepted or rejected for enrolment in a professional school. And, for students completing a Masters or Honours degree the "First Class Honours" label may be the admission ticket to an academic career or a post-graduate scholarship. While the attack on grading has not yet reached New Zealand universities to any significant degree it might be of interest to note the kind of arguments proposed by the advocates of grade abolition overseas.

36. At the outset those who seek to abolish grades might argue, in a variety of ways, that there are no valid and reliable tests of academic ability. Some basis for this point of view can be inferred from the known fact that the performances on standard I.Q. and achievement tests of many groups of people of non-European origin are consistently lower than for those of European origin. The argument goes on to assert that if such tests were culturally unbiassed the scores of distinguishable racial groups would be less varied. In the New Zealand context it would be argued that Maori under-representation at university level might be countered by staging a fundamental attack on the whole academic assumption of basing rewards on culturally-biased tests of ability. A second prong of an attack on grades could derive from the suggestion that academic ability as at present tested is only one of a number of abilities that should be valued by society. The qualities of leadership needed by those who administer and lead organisations, are not qualities that are readily related to the results of current assessment procedures, yet they are important qualities needed by the kinds of people that universities should be producing. This aspect of the argument against grades finds considerable backing in the results of studies which examine the relationship between university grades and adult ach'evement. One major review study (1) looked at forty-six different studies dealing with career categories, such as business, teaching, engineering, medicine and scientific research

(1) Donald P. Hoyt "The relationship between college grades and adult achievement: a review of the literature," American College Testing Program, Research Reports No. 7, September 1965.

and concluded that

> "although this area of research is plagued by many theoretical, experimental, measurement, and statistical difficulties, present evidence strongly suggests that college grades bear little or no relationship to any measure of adult accomplishment."

Such a conclusion, hitting so deeply as it does at the very basis of university staff recruitment patterns, will doubtless strike few chords of sympathy among academics themselves. For, to those who have achieved their own positions by demonstrating their academic ability, and who accept without question that such ability can or should be a fair basis for reward distribution, the arguments against grades pose personally disturbing questions.

37.. Those who, willingly or reluctantly, endorse the grading of students may well accept some of the views expressed by their opponents. They may agree that grading is neither accurate nor uniform, that it does not really standardize academic achievement criteria, that it produces undue competitive strain, that it protects bad teaching, or that it fails to present an accurate model of "real life." They may agree that grades frequently have a low correlation with creative achievement and that a motivation to earn a high grade is not identical to a motivation to learn. At the same time, they could well argue that academic standards, however incapable of clear definition, must be maintained because human beings are not capable of accepting a real life situation where all rewards become meaningless because they are distributed at random. Some cut-off point must be used to enable people to be sure, for example, that their doctor is not simply an undeserving recipient of automatic certification.

38. At this point the advocates of grade abolition might re-enter the argument by suggesting that they seek a more moderate reform - the introduction of pass/fail grading. While this proposal does not cover the problem of how to determine the pass/fail line it does entail the view that, for those who do pass, the existence of grades of passes is, in some way, irrelevant to the assessment of students. For example, while the broad categories of competent/ incompetent might be justified there are no sound reasons for assuming that a student four C passes is superior to a student with three C passes and a D failure. Similarly a student with B,C,D,D gradings in four subjects might be superior to a student with AAEE gradings even though both have two passes and the latter student has the better passes. Whether argumentation of this type should prevail over contentions that graded evaluation is a necessary concommitant of good teaching is a question that we shall, in this paper, bypass. Academics might well ask themselves, however, why they allocate grades to students, how they establish their standards, what distinghises a poor B grade for a good C grade, and whether grading in any way improves the quality of student learning.

Restricted and Aegrotat Passes

39. There are a number of devices that can be introduced into an examination system to ensure that students are able to have a fairer chance than if their sole prospects are through end-of-year examinations. None of these devices are necessarily relevant to a system of continuous assessment but can and have been used in New Zealand universities in particular circumstances in conjunction with the normal examination system. In essence they involve modifications of examination results rather than changes in assessment procedures and are therefore mentioned in this paper only briefly and descriptively.

40. A restricted pass may be given to a student whose examination results make him a marginal failure. Such a pass could be credited towards the degree results but would not enable a student to enter the next stage of the subject. This approach would enable a department to formally recognise a reasonable performance by a student who was unlikely to successfully complete a higher stage of study. With suitable limitations on the number of restricted passes allowable in any one degree some students might be able to complete their study more quickly.

41. Supplementary Examinations have been held in some New Zealand universities for some time for students in a number of professional courses in which they must pass the year's work as a whole to progress to a higher level. In the general arts and science degrees such examination might enable some students to upgrade poor passes or to reduce the total time required for graduation by passing an examination which has previously been failed. In fairness it should be said that those New Zealand universities that have considered running supplementary examinations in some general degree subjects have raised plausible objections - that costs would be too high, staff research could suffer, students in full time vacation work might not be able to adequately prepare for "specials," and that where the year's work is already being considered there is no need for supplementary examinations. However, while there may be little purpose served by an extension of the current form of written supplementary examinations held in late January as an examination rerun there could be justification for supplementary oral examinations. Notwithstanding the limitations of oral examining they may be used to supplement written examinations by enabling a marginal candidate to elaborate on insufficient answers, to show he has an adequate knowledge of the course it has answered too few questions to explain an unusual discrepancy in the standard of his answers, to correct what appears to be an accidental error, or to develop a promising argument. Supplementary oral examinations with these aims could be used with marginal students for the purposes attributed to written supplementary examinations. In the Victoria University Report already referred to, oral examinations were in fact discussed under the heading "reconsideration of scripts" (paras 56-61). That report rightly cites research evidence on the unreliability of oral examining but fails to point out that the bulk of the evidence available relates to research which contrasts written with oral examinations rather than written examinations with written examinations supplemented by oral examinations

for marginal candidates. While, then, oral examinations alone are of doubtful value one cannot conclude that they would necessarily be of little value in a supplementary rather than dominant position. An examiner could, in going over a script with a candidate, invite him to write a paragraph or two if he is experiencing difficulty in oral formulation. This approach could diminish the problem of the alleged unfamiliarity of New Zealand students with oral exposition. It would seem, in fact, that objections to oral examinations for marginal candidates derive more from intuition than from evidence. Their greatest demonstrable drawback is, in fact, the imposition of more responsibility on the examiner.

Conclusion

42. Formal examinations remain the mainstay of much assessment in the university. They are defended vigorously by some staff members, possibly those who have relied most heavily on them over a period of years and have not ventured into other assessment techniques. The comments in this paper may well seem to such academics to be a threat to cherished traditions. To other academics, particularly those with an interest in, and understanding of, the wealth of experience of those whose experimentation goes far beyond the conservative confines of this paper, my analysis and comments may seem quite archaic. Hopefully, those at both extremes may have found some comments of relevance and interest. For, if "continuous assessment" in some form or other is not yet acceptable to our university teachers, they may still be able to attempt minor and conservative reforms of the type hinted at in this paper.

APPENDIX I

The Efficient Use of Laboratory Time in the Teaching of Engineering

Mr D.E.P. Jenkins.

Traditionally laboratory experiments are used as a means of supplementing the lecture and the tutorial, and experiments are performed in order to verify certain aspects of the theory, If the student attends the laboratory regularly and writes laboratory reports of the required standard, then this is often considered to be efficient use of laboratory time.

The difficulty arises if one wants to deviate from this pattern. Indeed, one has to make a positive effort even to think about a change in the pattern, and when one has reached this stage it becomes necessary to ask some questions. Firstly, why is the change necessary? Secondly, what changes are to be made? Will the change be an improvement, and what benefits will the student enjoy from any changes that are made?

It is interesting to look quickly at the well-known pattern of a great deal of laboratory work in science and engineering. The student works from a laboratory sheet on which the main aspects of the experiments are outlined, and he is expected to carry out work according to the general directions of the sheets and draw sensible conclusions from his results. This is an all too familiar pattern of which most of us have experience. The main criticism, I feel, of this system is that it does not encourage creative thinking. This has been recognised for a very long time and some attempts have been made in

The Efficient Use of Laboratory Time in the Teaching of Engineering

Mr D.E.P. Jenkins.

Traditionally laboratory experiments are used as a means of supplementing the lecture and the tutorial, and experiments are performed in order to verify certain aspects of the theory, If the student attends the laboratory regularly and writes laboratory reports of the required standard, then this is often considered to be efficient use of laboratory time.

The difficulty arises if one wants to deviate from this pattern. Indeed, one has to make a positive effort even to think about a change in the pattern, and when one has reached this stage it becomes necessary to ask some questions. Firstly, why is the change necessary? Secondly, what changes are to be made? Will the change be an improvement, and what benefits will the student enjoy from any changes that are made?

It is interesting to look quickly at the well-known pattern of a great deal of laboratory work in science and engineering. The student works from a laboratory sheet on which the main aspects of the experiments are outlined, and he is expected to carry out work according to the general directions of the sheets and draw sensible conclusions from his results. This is an all too familiar pattern of which most of us have experience. The main criticism, I feel, of this system is that it does not encourage creative thinking. This has been recognised for a very long time and some attempts have been made in engineering to overcome the problem, notably by the introduction of simple projects in the early years of the course. Although this is a useful tool it does not in my view provide an entirely satisfactory solution; the whole problem is much more fundamental, as the only satisfactory way of dealing with laboratory work is to treat it as an integral part of the course, of equal importance to the lecture and the tutorial.

Supposing one takes the following standpoint: (a) one has three or four years to impart information to the student; (b) the information to be conveyed is determined by the syllabus; (c) in order to enable the lecturers to undertake the task the university provides the following tools: lecturing facilities, tutorial facilities, library facilities and laboratory facilities.

The next stage in the analysis is to look at the subject as a whole and determine which parts of the course are best done by lecture, which by tutorials or discussions, which by private reading and searching in the library, and which in the laboratory. This is a very difficult task and is one which in my experience many lecturers are not prepared to undertake.

Having analysed the course as a whole and determined the topics which are best taught by experimental techniques, we are in a position to ask how we should use the experimental techniques in relation to the selected topics, and what latent abilities in the student we wish to develop in the time devoted to the experimental techniques. We should attempt to set down some

of the objectives of a scheme of laboratory work. For example, that it should promote a true spirit of inquiry and ordered thinking, provide maximum student involvement and hence motivation, demand the active participation of all students in both planning and performing experimental work, encourage students to criticise and accept criticism, provide an environment where the maximum level and standard of work is limited only by the students' ability, enthusiasm and the time available, develop in the case of engineering a sound basis for tackling real engineering situations, in particular concepts of engineering compromise, provide a forum for general discussion and teach the correct use of library facilities and other reference sources.

All this you have doubtless heard many times before and it may sound rather hackneyed, but if you accept these as being realistic aims of any teaching situation and as the basis of planning a programme of experimental work you are immediately faced with a very challenging problem. I am now in a position to describe how we have faced this challenge in the Electrical Engineering Department of the City University, both at the first year and third year levels of a four year course.

Let us first look at the work at first year level. Each week the students receive two fifty-minute lectures of the conventional type, but confined to selected topics, mainly fundamental; then one large group tutorial of about thirty students, when all the lecturers concerned with the work are present for discussion and problem solving; then three hours allocated to experimental engineering, as the non-lecturing aspects are called. For the whole of this three-hour period the laboratory is available for use by the students.

At the beginning of the term the student is presented with a handout consisting of several sheets of carefully prepared questions, which range from the apparently very simple to the very difficult. There are two more such handouts during the term, and each is related to broad areas of work, conveniently grouped in this particular subject to direct current theory, alternating current theory and electro-magnetic field theory. The student is expected to answer as many of these questions as he is able, using the tools at his disposal. Some of the questions can be answered by reference to the theory, others can only be solved by carrying out experimental work. The student himself must decide how to tackle the problems. If he decides to carry out an experiment he must first of all design it, then select the equipment, carry out the work and produce the results. The answers to all the questions, theoretical and practical, are written in a laboratory log book. In this way he builds up a complete picture of information on the particular subject of the syllabus. There is no formal written-up report on the experimental work, and the log book is the only record of the student's effort. This is corrected by the lecturer-tutor with a view to commenting on the work to help the student to understand the subject and recognise his own deficiencies.

Let us now think for a moment about the supervision which is associated with this kind of project. Four lecturers as a group are responsible for the first-year instruction of a student in this particular subject. The intake at the moment is about sixty or seventy and, as you may know, in our particular

institution we run a sandwich system which has two entries per year, one in September and one in February, both of which are about sixty or seventy. Each of the lecturers, therefore, acts as a tutor to about fifteen or eighteen students.

The students work in groups of three for the experimental engineering periods, three such groups per lecturer being in the laboratory during any one three-hour period. Approximately six weeks are devoted to each set of questions, the laboratory being available on alternate weeks at set times, but the students are permitted to use it at any other time provided there is adequate supervision. Each set of questions is introduced by the lecturer in a discussion period, and similarly summarised and discussed in another period. The lecturer tutor is available in the laboratory for the whole of the three hours, and normally deals only with his own group of students. In practice it is a continuous process of discussion, questions and answers, with on the spot demonstrations of difficult topics and analyses of incorrect procedures. Often the students are taken to a nearby tutorial room for short discussions on important issues which arise from the experimental work. In this way the lecturer gets to know his students very well, and he is able to guide the students, working on an individual basis.

The pattern of three students per group using the laboratory on alternate weeks is determined entirely by space considerations and availability of equipment. Ideally we think the student should work independently and associated with and adjacent to the laboratory there should be tutorial rooms and a departmental library, so that the students use the laboratory as a base for their studies.

In the third-year work, where at the moment the

institution we run a sandwich system which has two entries per year, one in September and one in February, both of which are about sixty or seventy. Each of the lecturers, therefore, acts as a tutor to about fifteen or eighteen students.

The students work in groups of three for the experimental engineering periods, three such groups per lecturer being in the laboratory during any one three-hour period. Approximately six weeks are devoted to each set of questions, the laboratory being available on alternate weeks at set times, but the students are permitted to use it at any other time provided there is adequate supervision. Each set of questions is introduced by the lecturer in a discussion period, and similarly summarised and discussed in another period. The lecturer tutor is available in the laboratory for the whole of the three hours, and normally deals only with his own group of students. In practice it is a continuous process of discussion, questions and answers, with on the spot demonstrations of difficult topics and analyses of incorrect procedures. Often the students are taken to a nearby tutorial room for short discussions on important issues which arise from the experimental work. In this way the lecturer gets to know his students very well, and he is able to guide the students, working on an individual basis.

The pattern of three students per group using the laboratory on alternate weeks is determined entirely by space considerations and availability of equipment. Ideally we think the student should work independently and associated with and adjacent to the laboratory there should be tutorial rooms and a departmental library, so that the students use the laboratory as a base for their studies.

In the third-year work, where at the moment the numbers are smaller because the sixty/seventy intake has not yet reached the third year, the approach has proceeded one stage further. Here the laboratory is actually used as a basis for teaching the subjects, in this case networks and power systems. The lecture is used only to explain fundamental concepts. The experiments in the laboratory are designed to involve the student in the use of these fundamental concepts and engineering practice and give him the opportunity to exercise his powers of critical thinking. Carefully constructed numerical and theoretical examples are given to the students and these are dealt with in the tutorial periods-cum-discussion groups of about six students with the lecturers present. Each student is given the responsibility for solving and understanding one example in each tutorial period. He has to provide his solution and explanation to his fellow students, and the lecturers who are present, by means of a short lecture/talk/ demonstration. The essence of this approach is again a very close relationship between the lecturers and the students, and also the use of the laboratory as a teaching tool and a base for study.

Let us now consider for a moment or two student/ staff reaction to these schemes, in particular in the first-year work. The immediate reaction of the student is that this approach is elementary and is a little beneath him. This is natural since the equipment they use is usually very simple and the experimental techniques unglamorous, and it is important to bear in mind that these students from good grammar schools will in their Upper VI physics work often have used very sophisticated equipment. This attitude soon changes to one of interest and

later immense enthusiasm, in our experience. This inevitably leads to a comparison with the laboratory work they carry out in other departments on more traditional lines, and there is no doubt in our view they see the value of the new approach we are using. The log book entries are variable; the good students usually produce very good work but the others have to be encouraged. Student participation is the striking feature of this work. They learn from one another and develop their ability to ask questions.

Now the lecturer reaction: this is extremely hard work and the laboratory sessions often become three hours of concentration and discussion and explanation - much more difficult than supervising set tasks in the laboratory. The preparation of the question sheets is probably the most difficult part of the whole exercise. They have to be carefully constructed and amended in the light of experience. The analysis and criticism of the log book has to be done thoroughly and conscientiously and discussed with the student if he is to learn from his mistakes.

In conclusion I would like to bring your thoughts to just four more points which in my view are most important. I would first of all like to echo Professor Tyrrell's comments and say that material presented in the early years of the work must be relevant and must be seen by the students to be relevant. This is very important. Secondly, a student who has been educated in a system such as I have described, a system based on an experimental approach to learning, has been trained to analyse, ask questions and generally use his powers of creativity and critical thinking. He will feel severely let down if in his next years of the subject he reverts to traditional laboratory techniques. This is a very serious problem indeed, since the second and third years of the course become increasingly difficult to deal with in this way. The final year is easier, since it is then possible to use the advanced project as a means of stimulating the student. These comments are, of course, also applicable to students moving into university life from the grammar school, particularly those students who are now coming up who have been taught on the Nuffield science project method, because the technique is exactly the same and this transfer becomes exactly the same.

Thirdly, in my experience lecturers are very conservative and resist change. A scheme such as the one I have outlined can meet with a great deal of opposition. Initially it must be carried through by one or two enthusiasts who by their demonstration of enthusiasm will win others in support of their ideas. Lastly, the student usually has no choice in the matter; he has to participate in the new schemes. Nevertheless, he must feel confident that his education will not suffer as a result of his participation in innovations. As educators we have a dual responsibility: one, we must make advances in teaching methods, but, two, we must also ensure an adequate education for the students in our charge.

From: Innovations and Experiments in University Teaching Methods, a report of the proceedings at the third conference organised by the University Teaching Methods Research Unit of the Department held at the Institute of Education, 3rd - 4th April, 1968, pps 25-29.